References

[1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57 (1):289–300, 1995.

[2] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*. Springer Science & Business Media, 2005.

[3] Seymour Gkisser. *Predictive inference: an introduction*. Chapman and Hall/CRC, 2017.

[4] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

[5] Bradley Efron, Robert Tibshirani, John D Storey, and Virginia Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American statistical association*, 96(456):1151–1160, 2001.

[6] Christopher Genovese and Larry Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):499–517, 2002.

[7] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[8] John D Storey, Jonathan E Taylor, and David Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

[9] Bradley Efron. Local false discovery rates, 2005.

[10] Adel Javanmard and Andrea Montanari. Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory. *IEEE Transactions on Information Theory*, 60 (10):6522–6554, 2014.

[11] Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085, 2015.

[12] Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection. *arXiv preprint arXiv:1610.02351*, 2016.

[13] Wenguang Sun and T Tony Cai. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.

[14] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[15] Jing Lei. Classification with confidence. *Biometrika*, 101(4):755–769, 2014.

[16] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523): 1094–1111, 2018.

[17] Yaniv Romano, Evan Patterson, and Emmanuel J Candès. Conformalized quantile regression. *arXiv preprint arXiv:1905.03222*, 2019.

[18] Anastasios Angelopoulos, Stephen Bates, Jitendra Malik, and Michael I Jordan. Uncertainty sets for image classifiers using conformal prediction. *arXiv preprint arXiv:2009.14193*, 2020.

[19] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

[20] Stephen J Roberts and Will D Penny. Variational bayes for generalized autoregressive models. *IEEE Transactions on Signal Processing*, 50(9):2245–2257, 2002.

[21] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[22] Percy Liang, Slav Petrov, Michael I Jordan, and Dan Klein. The infinite pcfg using hierarchical dirichlet processes. In *Proceedings of the 2007 joint conference on empirical methods in natural*

language processing and computational natural language learning (EMNLP-CoNLL), pages 688–697, 2007.

[23] Guido Sanguinetti, Neil D Lawrence, and Magnus Rattray. Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics*, 22(22):2775–2781, 2006.

[24] Kenichi Kurihara and Taisuke Sato. Variational bayesian grammar induction for natural language. In *International Colloquium on Grammatical Inference*, pages 84–96. Springer, 2006.

[25] Peter Carbonetto and Matthew Stephens. Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies. *Bayesian Analysis*, 7(1):73–108, 2012.

[26] Anil Raj, Matthew Stephens, and Jonathan K Pritchard. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, 197(2):573–589, 2014.

[27] Nebojsa Jojic and Brendan J Frey. Learning flexible sprites in video layers. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[28] Michael Braun and Jon McAuliffe. Variational inference for large-scale models of discrete choice. *Journal of the American Statistical Association*, 105(489):324–335, 2010.

[29] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of 'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

[30] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[31] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[32] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[33] Antonio Auffinger, Gérard Ben Arous, and Jiří Černỳ. Random matrices and complexity of spin glasses. *Communications on Pure and Applied Mathematics*, 66(2):165–201, 2013.

[34] Noureddine El Karoui, Derek Bean, Peter J Bickel, Chinghway Lim, and Bin Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

[35] Noureddine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results. *arXiv preprint arXiv:1311.2445*, 2013.

[36] Mihailo Stojnic. A framework to characterize performance of lasso algorithms. *arXiv preprint arXiv:1303.7291*, 2013.

[37] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi. Regularized linear regression: A precise analysis of the estimation error. In *Conference on Learning Theory*, pages 1683–1709. PMLR, 2015.

[38] David Donoho and Andrea Montanari. High dimensional robust m-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3):935–969, 2016.

[39] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi. Precise error analysis of regularized M-estimators in high dimensions. *IEEE Transactions on Information Theory*, 64(8):5592–5628, 2018.

[40] Xiaoyi Mai, Zhenyu Liao, and Romain Couillet. A large scale analysis of logistic regression: Asymptotic performance and new insights. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3357–3361. IEEE, 2019.

[41] Pragya Sur and Emmanuel J Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525, 2019.

[42] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1): 33–79, 2020.

[43] Emmanuel J Candès, Pragya Sur, et al. The phase transition for the existence of the maximum likelihood estimate in high-dimensional logistic regression. *The Annals of Statistics*, 48(1):27–42, 2020.

[44] Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019.

[45] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1):1–45, 2021.

[46] David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[47] Judea Pearl. Reverend bayes on inference engines: a distributed hierarchical approach. In *Proceedings of the Second AAAI Conference on Artificial Intelligence*, pages 133–136, 1982.

[48] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.

[49] Thomas Peter Minka. *A family of algorithms for approximate Bayesian inference*. PhD thesis, Massachusetts Institute of Technology, 2001.

[50] Peter Hall, John T Ormerod, and Matt P Wand. Theory of Gaussian variational approximation for a Poisson mixed model. *Statistica Sinica*, pages 369–389, 2011.

[51] Peter Hall, Tung Pham, Matt P Wand, and Shen SJ Wang. Asymptotic normality and valid inference for Gaussian variational approximation. *The Annals of Statistics*, 39(5):2502–2532, 2011.

[52] Peter Bickel, David Choi, Xiangyu Chang, and Hai Zhang. Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *Annals of Statistics*, 41(4): 1922–1943, 2013.

[53] Chong You, John T Ormerod, and Samuel Mueller. On variational Bayes estimation and variational information criteria for linear regression models. *Australian & New Zealand Journal of Statistics*, 56 (1):73–87, 2014.

[54] Anderson Y Zhang and Harrison H Zhou. Theoretical and computational guarantees of mean field variational inference for community detection. *Annals of Statistics*, 48(5):2575–2598, 2020.

[55] Behrooz Ghorbani, Hamid Javadi, and Andrea Montanari. An instability in variational inference for topic models. In *International conference on machine learning*, pages 2221–2231. PMLR, 2019.

[56] Yixin Wang and David M Blei. Frequentist consistency of variational Bayes. *Journal of the American Statistical Association*, 114(527):1147–1161, 2019.

[57] Sourav Chatterjee and Amir Dembo. Nonlinear large deviations. *Advances in Mathematics*, 299: 396–450, 2016.

[58] Fengshuo Zhang and Chao Gao. Convergence rates of variational posterior distributions. *Annals of Statistics*, 48(4):2180–2207, 2020.

[59] Sumit Mukherjee and Subhabrata Sen. Variational inference in high-dimensional linear regression. *arXiv preprint arXiv:2104.12232*, 2021.

[60] Peter Müller, Giovanni Parmigiani, Christian Robert, and Judith Rousseau. Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of the American Statistical Association*, 99(468):990–1001, 2004.

[61] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.

[62] Martin A Lindquist and Amanda Mejia. Zen and the art of multiple comparisons. *Psychosomatic medicine*, 77(2):114, 2015.

[63] Yoav Benjamini and Rami Cohen. Weighted false discovery rate controlling procedures for clinical trials. *Biostatistics*, 18(1):91–104, 2017.

[64] William Fithian and Lihua Lei. Conditional calibration for false discovery rate control under dependence. *arXiv preprint arXiv:2007.10438*, 2020.

[65] Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values. *arXiv preprint arXiv:2009.02824*, 2020.

[66] Molei Liu, Eugene Katsevich, Lucas Janson, and Aaditya Ramdas. Fast and powerful conditional randomization testing via distillation. *arXiv preprint arXiv:2006.03980*, 2020.

[67] Asaf Weinstein, Rina Barber, and Emmanuel Candes. A power and prediction analysis for knockoffs with lasso statistics. *arXiv preprint arXiv:1712.06465*, 2017.

[68] Asaf Weinstein, Weijie J Su, Malgorzata Bogdan, Rina F Barber, and Emmanuel J Candes. A power analysis for knockoffs with the lasso coefficient-difference statistic. *arXiv preprint arXiv:2007.15346*, 2020.

[69] Wenshuo Wang and Lucas Janson. A power analysis of the conditional randomization test and knockoffs. *arXiv preprint arXiv:2010.02304*, 2020.

[70] Peter Hoff. Bayes-optimal prediction with frequentist coverage control. *arXiv preprint arXiv:2105.14045*, 2021.

[71] Scott Kirkpatrick and David Sherrington. Infinite-ranged models of spin-glasses. *Physical Review B*, 17(11):4384, 1978.

[72] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[73] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3):859–929, 2019.

[74] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability theory and related fields*, 174(3):1133–1185, 2019.

[75] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

[76] Michael Celentano, Zhou Fan, and Song Mei. Local convexity of the tap free energy and amp convergence for z2-synchronization. *arXiv preprint arXiv:2106.11428*, 2021.

[77] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

[78] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[79] Sandrine Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.

[80] Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, et al. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14(5):566–806, 2021.

[81] Timm Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971, 1982.

[82] Cyrano De Dominicis and A. Peter Young. Weighted averages and order parameters for the infinite range Ising spin glass. *Journal of Physics A: Mathematical and General*, 16(9):2063, 1983.

[83] AJ Bray, Michael A. Moore, and A. Peter Young. Weighted averages of TAP solutions and parisi's $q(x)$. *Journal of Physics C: Solid State Physics*, 17(5):L155, 1984.

[84] Andrea Cavagna, Irene Giardina, Giorgio Parisi, and Marc Mézard. On the formal equivalence of the TAP and thermodynamic methods in the SK model. *Journal of Physics A: Mathematical and General*, 36(5):1175, 2003.

[85] Sourav Chatterjee. Spin glasses and Stein's method. *Probability theory and related fields*, 148(3-4): 567–600, 2010.

[86] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

[87] Antonio Auffinger and Aukosh Jagannath. Thouless-Anderson-Palmer equations for conditional Gibbs measures in the generic p-spin glass model. *arXiv preprint arXiv:1612.06359*, 2016.

[88] Wei-Kuo Chen and Dmitry Panchenko. On the TAP free energy in the mixed p-spin models. arXiv:1709.03468, 2017.

[89] Wei-Kuo Chen, Dmitry Panchenko, and Eliran Subag. The generalized TAP free energy. *arXiv preprint arXiv:1812.05066*, 2018.

[90] David Belius and Nicola Kistler. The TAP–Plefka variational principle for the spherical SK model. *Communications in Mathematical Physics*, 367(3):991–1017, 2019.

[91] Eliran Subag. The free energy of spherical pure p-spin models–computation from the TAP approach. *arXiv preprint arXiv:2101.04352*, 2021.

[92] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.

[93] Rong Ge, Jason D Lee, and Tengyu Ma. Matrix completion has no spurious local minimum. *arXiv preprint arXiv:1605.07272*, 2016.

[94] Song Mei, Yu Bai, Andrea Montanari, et al. The landscape of empirical risk for nonconvex losses. *Annals of Statistics*, 46(6A):2747–2774, 2018.

[95] Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, 176(1):5–37, 2019.

[96] Amit Singer and Hau-tieng Wu. Orientability and diffusion maps. *Applied and computational harmonic analysis*, 31(1):44–58, 2011.

[97] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2): 125–170, 2016.

[98] Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276, 1998.

[99] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

[100] Cynthia Rush and Ramji Venkataramanan. Finite sample analysis of approximate message passing algorithms. *IEEE Transactions on Information Theory*, 64(11):7264–7286, 2018.

[101] Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, Lenka Zdeborová, et al. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems*, 29:424–432, 2016.

[102] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method. *arXiv preprint arXiv:1812.02537*, 2018.

[103] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321–345, 2021.

[104] Xinyi Zhong, Chang Su, and Zhou Fan. Empirical bayes pca in high dimensions. *arXiv preprint arXiv:2012.11676*, 2020.

[105] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Neural Information Processing Systems*, 2016.

[106] Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory*, 66(7):4270–4303, 2020.

[107] Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. High-temperature expansions and message passing algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(11):113301, 2019.

[108] Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.

[109] Michael Celentano, Andrea Montanari, and Yuting Wei. The Lasso with general Gaussian designs with applications to hypothesis testing. *arXiv preprint arXiv:2007.13716*, 2020.

[110] Xin Xing, Zhigen Zhao, and Jun S Liu. Controlling false discovery rate using gaussian mirrors. *Journal of the American Statistical Association*, (just-accepted):1–45, 2021.

[111] Chenguang Dai, Buyu Lin, Xin Xing, and Jun S Liu. A scale-free approach for false discovery rate control in generalized linear models. *arXiv preprint arXiv:2007.01237*, 2020.

[112] David John Cameron Mackay. *Bayesian methods for adaptive models*. PhD thesis, California Institute of Technology, 1992.

[113] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[114] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *arXiv preprint arXiv:1703.04977*, 2017.

[115] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. *arXiv preprint arXiv:1802.10501*, 2018.

[116] Wesley J Maddox, Pavel Izmailov, Timur Garipov, Dmitry P Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. *Advances in Neural Information Processing Systems*, 32:13153–13164, 2019.

[117] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*, 2016.

[118] Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, David Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. *arXiv preprint arXiv:1906.02530*, 2019.

[119] Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E Hopcroft, and Kilian Q Weinberger. Snapshot ensembles: Train 1, get m for free. *arXiv preprint arXiv:1704.00109*, 2017.

[120] Andrey Malinin, Bruno Mlodozeniec, and Mark Gales. Ensemble distribution distillation. *arXiv preprint arXiv:1905.00076*, 2019.

[121] Anastasios N Angelopoulos, Stephen Bates, Emmanuel J Candès, Michael I Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control. *arXiv preprint arXiv:2110.01052*, 2021.

[122] Song Mei and Andrea Montanari. The generalization error of random features regression: Precise asymptotics and the double descent curve. *Communications on Pure and Applied Mathematics*, 2019.

[123] Zitong Yang, Yaodong Yu, Chong You, Jacob Steinhardt, and Yi Ma. Rethinking bias-variance trade-off for generalization of neural networks. In *International Conference on Machine Learning*, pages 10767–10777. PMLR, 2020.

[124] Zitong Yang, Yu Bai, and Song Mei. Exact gap between generalization error and uniform convergence in random feature models. *arXiv preprint arXiv:2103.04554*, 2021.

[125] Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Generalisation error in learning with random features and the hidden manifold model. In *International Conference on Machine Learning*, pages 3452–3462. PMLR, 2020.